(19)日本国特許庁（ＪＰ）　(12) 公開特許公報（A）　(11)特許出願公開番号

特開平11－249679

(43)公開日　平成11年(1999)９月17日

(51)Int.Cl.[6]		識別記号	ＦＩ		
Ｇ１０Ｌ	3/00		Ｇ１０Ｌ	3/00	Ｈ
	5/04			5/04	Ｆ

審査請求　未請求　請求項の数５　ＯＬ　（全 ５ 頁）

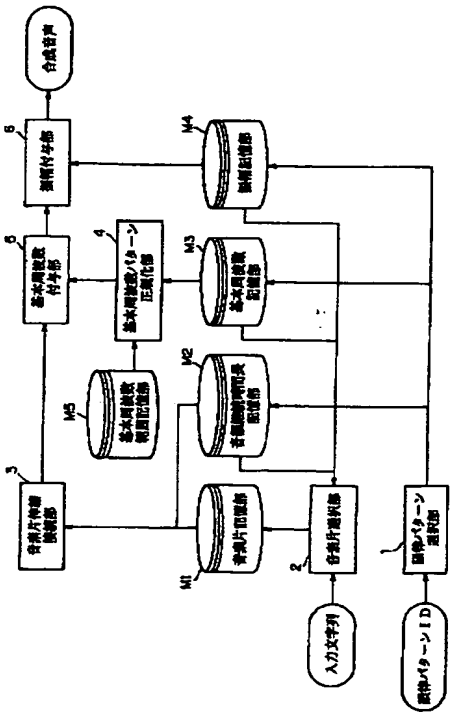
(21)出願番号　特願平10－52361

(22)出願日　平成10年(1998)３月４日

(71)出願人　000006747
株式会社リコー
東京都大田区中馬込１丁目３番６号
(72)発明者　酒寄　哲也
東京都大田区中馬込１丁目３番６号　株式会社リコー内
(74)代理人　弁理士　高野　明近

(54)【発明の名称】　音声合成装置

(57)【要約】
【課題】　入力された文字列から定型的なフレーズを合成音声で読み上げる際の合成音声を、より自然性の音声に近いものにすること。
【解決手段】　人間が発声したフレーズから抽出した音韻継続時間長系列、基本周波数系列、振幅あるいはパワーの系列をそれぞれ記憶した、音韻継続時間記憶部、基本周波数記憶部、振幅記憶部を備え、かつ、音素あるいは音素連鎖を音韻単位とし音韻情報を音素片としてを記憶する音素片記憶部を備えた音声合成装置により、入力文字列にしたがって音素片記憶部から読み出した音素片系列を、韻律パターン記憶部から読み出した音韻継続時間長系列にしたがって伸縮して接続し、韻律パターン記憶部から読み出した基本周波数系列にしたがって基本周波数付与を行い、その際、基本周波数範囲内に収まるように正規化を施し、更に、韻律パターン記憶部から読み出した振幅又はパワー系列にしたがって振幅付与を行って音声を合成する。



1

【特許請求の範囲】

【請求項１】　人間が発声したフレーズから抽出した音韻継続時間長系列、基本周波数系列、振幅あるいはパワーの系列をそれぞれ記憶した、音韻継続時間記憶部、基本周波数記憶部、振幅記憶部、及び、音素あるいは音素連鎖を音韻単位とし音韻情報を音素片として記憶する音素片記憶部を具備し、入力文字列にしたがって音素片記憶部から読み出した音素片系列を、韻律パターン記憶部から読み出した音韻継続時間長系列にしたがって伸縮して接続し、韻律パターン記憶部から読み出した基本周波数系列にしたがって基本周波数付与を行い、韻律パターン記憶部から読み出した振幅又はパワー系列にしたがって振幅付与を行って音声を合成することを特徴とする音声合成装置。

【請求項２】　請求項１に記載された音声合成装置において、音素片記憶部に記憶した音素片セットによって無理なく合成できる基本周波数の範囲を記憶する基本周波数範囲記憶部を具備し、基本周波数記憶部から読み出した基本周波数系列に対して、基本周波数範囲記憶部から読み出した基本周波数範囲内に収まるように正規化を施して音声を合成することを特徴とする音声合成装置。

【請求項３】　請求項１に記載された音声合成装置において、前記音素片記憶部は一つの音韻単位に対して適応すべき基本周波数範囲毎に複数の音素片を記憶しており、前記基本周波数記憶部から読み出した基本周波数に対応した音素片を選択的に用いて音声合成を行うことを特徴とする音声合成装置。

【請求項４】　請求項１に記載された音声合成装置において、前記音素片記憶部は一つの音韻単位に対して適応すべき振幅範囲毎に複数の音素片を記憶しており、基本周波数記憶部から読み出した振幅あるいはパワーに対応した音素片を選択的に用いて音声合成を行うことを特徴とする音声合成装置。

【請求項５】　請求項１に記載された音声合成装置において、前記音素片記憶部は一つの音韻単位に対して適応すべき音韻継続時間長範囲毎に複数の音素片を記憶しており、音韻継続時間長記憶部から読み出した音韻継続時間長に対応した音素片を選択的に用いて音声合成を行うことを特徴とする音声合成装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、定型文章と非定型文章の混在するテキストを音声に変換するような用途に利用し得る音声合成装置に関するものである。

【０００２】

【従来の技術】特開平５－２７７８９号公報には、「音声合成装置」が記載されている。これは、規則合成、分析合成、録音再生など異なる方式の合成音声を混在して用いる場合に、接続部分でオーバーラップ処理を行うものである。特開平８－６３１８７号公報には、定型部分

2

に対して自然音声から抽出した基本周波数と音韻時間長を用いる「音声合成装置」が開示されている。

【０００３】

【発明が解決しようとする課題】従来、音声合成装置の合成方式には録音編集方式と規則合成方式がある。前者はアナウンサーなどがフレーズ毎に音声を登録しておき、これを適宜選択結合してメッセージを作成するもので、肉声に近い良好な音声が得られる可能性がある反面、データ量が多い、登録外のフレーズには対応できない、新たにフレーズを追加するために同一話者の確保が必要であるなどの問題がある。他方、後者は音素や音節などの細かい単位で音声データを蓄積して任意語彙の合成を可能とするものであるが、音質的に録音編集方式に劣り、特に基本周波数、音韻継続時間長、振幅などの韻律パターンを規則によって付与するためどうしても機械的で不自然なものになる。

【０００４】このため任意語彙の出力が不要な定型的なメッセージには音質の良い録音編集方式が用いられ、テキストからの音声変換が必要な場面では規則合成が用いられる。しかし、カーナビゲーションの音声案内で定型的メッセージの中に地名が埋め込まれるなど、定型文の中に一部任意語彙が埋め込まれるようなアプリケーションも多く存在する。このような場合、ごく一部の任意語彙のために全体の音質を落として規則合成を採用するか、任意語彙の出力を諦めて録音編集方式を用いるか、あるいは定型部分を録音編集で行い任意語彙部分のみ規則合成で行うという混在方式を採るかの選択をせざるを得ない。

【０００５】録音編集と規則合成を混在させる場合の問題点は、２つの方式で出力音声の声質がまったく異なるため、聞いていて違和感があるばかりでなく非常に聞取り難いものとなる点である。前記特開平５－２７７８９号公報に記載された「音声合成装置」ではこの問題に対し、異なる方式間の出力音声をオーバーラップさせて接続することで対処している。しかし、このようにしても定型部分と任意語部分で話者が変わってしまうことは避けられず基本的な問題は解決していない。また、オーバーラップ部分では２人の話者が同時に話しているようになるため聞取り難くなる可能性がある。

【０００６】これに対し、前記特開平８－６３１８７号公報に記載された「音声合成装置」では、定型文にも規則合成的に音素あるいは音節等をつないで音韻パラメータを生成し、これに自然音声から抽出した基本周波数及び音韻継続時間長を付与することにより、任意語部分との話者連続性を保持しつつ自然性を向上している。しかし、様々な韻律パラメータや音韻パラメータは相互に関連があり、全体としてバランスを取るように構築されている規則群の一部（基本周波数と音韻継続時間長）だけを全く異なる話者特性、発声様式の音声から移植することは思わぬ不整合を生んで全体の自然性を損なう可能性

がある。例えば、文末にかけて基本周波数は大きく下がることがあるが、この時は振幅も十分小さくしないと不自然に低い声が目立つことになる。また、このような場合、本来口の開きも小さくなり音声スペクトル自体の変化があるはずであり、あまりに明瞭な音素片データを用いることも違和感を生む。さらに音素片データには対応可能な基本周波数の範囲が存在し、自然音声の基本周波数パターンはこれよりもダイナミックレンジが広いのが普通であるため、無理な基本周波数付与により明瞭性の低下を招く恐れがある。そこで、本発明はこのような問題点を解決し、定型的フレーズの合成音声の自然性を向上することを目的とする。

【０００７】

【課題を解決するための手段】請求項１の発明は、人間が発声したフレーズから抽出した音韻継続時間長系列、基本周波数系列、振幅あるいはパワーの系列をそれぞれ記憶した、音韻継続時間記憶部、基本周波数記憶部、振幅記憶部、及び、音素あるいは音素連鎖を音韻単位とし音韻情報を音素片として記憶する音素片記憶部を具備し、入力文字列にしたがって音素片記憶部から読み出し並べた音素片系列を、韻律パターン記憶部から読み出した音韻継続時間長系列にしたがって伸縮して接続し、韻律パターン記憶部から読み出した基本周波数系列にしたがって基本周波数付与を行い、韻律パターン記憶部から読み出した振幅又はパワー系列にしたがって振幅付与を行って音声を合成する音声合成装置である。

【０００８】請求項２の発明は、請求項１に記載された音声合成装置において、音素片記憶部に記憶した音素片セットによって無理なく合成できる基本周波数の範囲を記憶する基本周波数範囲記憶部を具備し、基本周波数記憶部から読み出した基本周波数系列に対して、基本周波数範囲記憶部から読み出した基本周波数範囲内に収まるように正規化を施して音声を合成する音声合成装置である。

【０００９】請求項３の発明は、請求項１に記載された音声合成装置において、前記音素片記憶部は一つの音韻単位に対して適応すべき基本周波数範囲毎に複数の音素片を記憶しており、前記基本周波数記憶部から読み出した基本周波数に対応した音素片を選択的に用いて音声合成を行う音声合成装置である。

【００１０】請求項４の発明は、請求項１に記載された音声合成装置において、前記音素片記憶部は一つの音韻単位に対して適応すべき振幅範囲毎に複数の音素片を記憶しており、基本周波数記憶部から読み出した振幅あるいはパワーに対応した音素片を選択的に用いて音声合成を行う音声合成装置である。

【００１１】請求項５の発明は、請求項１に記載された音声合成装置において、前記音素片記憶部は一つの音韻単位に対して適応すべき音韻継続時間長範囲毎に複数の音素片を記憶しており、音韻継続時間長記憶部から読み出した音韻継続時間長に対応した音素片を選択的に用いて音声合成を行う音声合成装置である。

【００１２】

【発明の実施の形態】本発明の音声合成装置の一実施例について説明する。図１は、この実施例における構成を示す。図１中、Ｍ１は音素あるいは音素連鎖を音韻単位とし音韻情報を音素片として記憶する音素片記憶部であって、一つの音韻単位に対して適応すべき基本周波数範囲毎に複数の音素片を記憶しており、基本周波数範囲記憶部Ｍ５から読み出した基本周波数に対応した音素片を選択的に用いる。また、Ｍ２，Ｍ３，Ｍ４はそれぞれ人間が発声したフレーズから抽出した音韻継続時間長系列、基本周波数系列、振幅あるいはパワーの系列をそれぞれ記憶した、音韻継続時間長記憶部、基本周波数記憶部、振幅記憶部を示している。Ｍ５は基本周波数範囲記憶部であって、音素片記憶部Ｍ１に記憶された音素片セットによって無理なく合成できる基本周波数の範囲を記憶している。

【００１３】各部の動作について以下に説明する。韻律パターン選択部１は入力される韻律パターン１Ｄ（韻律パターンを識別する識別子：例えば、番号等によりそれに対応するパターンを識別するもの）から音韻継続時間長、基本周波数、振幅の各パターンを選択する。音素片選択部２は入力文字列から音素片ラベルを得、また韻律パターン選択部１で選択された音韻継続時間長、基本周波数、振幅の各韻律パターンの範囲を参考にして、これらの情報を元に音素片記憶部Ｍ１から必要な音素片を検索する。

【００１４】図２は、音素片記憶部Ｍ１のデータ構造の一例を示したものである。同一音素ラベル、例えば、“ア”に対して異なる適用可能韻律パラメータ範囲のデータを複数記憶している。ここに示すように韻律パラメータ範囲は、例えば、時間長範囲の長短、増幅範囲の大小などカテゴライズされたものでも、基本周波数範囲のように下限値と上限値を示すものでもよく、またそれらが混在していても構わない。表中のデータ欄には実際には音素片データ、波形データ及びスペクトルパラメータが格納される。この中からラベルの一致するもので韻律パラメータ範囲が最も近いものを選択する。

【００１５】音素片伸縮接続部３は、入力文字列にしたがって音素片選択部２により選択された音素片系列を、韻律パターン選択部１で選択された音韻継続時間長の範囲を参考にして、音韻継続時間長記憶部Ｍ２から選択された音韻継続時間長パターンに従って伸縮してそれぞれの音素片を接続する。基本周波数範囲記憶部Ｍ５には、音素片記憶部Ｍ１に記憶された音素片データセット全体によってカバーされる基本周波数範囲が記憶されており、基本周波数パターン正規化部４は、基本周波数範囲記憶部Ｍ５から選択された基本周波数パターンがこの範囲を逸脱している場合に、選択された基本周波数パター

5

ンをこの範囲に合わせて正規化する。基本周波数付与部５は、音素片伸縮接続部３で接続された音素片系列パターンに対して、正規化された基本周波数パターンを付与する。

【００１６】振幅付与部６は接続されかつ基本周波数が付与された音素片系列パターンに対し、韻律パターン選択部１で選択された振幅パターンの範囲を参考にして、振幅記憶部Ｍ４から選択された振幅パターンを付与して合成音声を作成する。なお、音素片の伸縮及び接続、基本周波数及び振幅の付与に関しては規則音声合成の一般的技術を用いることが出来るため、ここでは詳細な説明は省略する。

【００１７】

【発明の効果】請求項１に対応する効果：基本的な韻律パラメータである基本周波数、音韻継続時間長、振幅の３つを同じ親善音声フレーズから抽出したものを使用することによって韻律パラメータ間の不整合を抑え、合成音声の自然性を向上することができる。

【００１８】請求項２に対応する効果：音素片データベースが対応可能な範囲に基本周波数を正規化することによって、無理な基本周波数付与を防ぎ合成音声の明瞭性の低下を防ぐことができる。

【００１９】請求項３に対応する効果：付与すべき基本周波数に対応する音素片データを選択的に用いることにより、ダイナミックレンジの広い自然音声の基本周波数を付与することが可能となり、明瞭性を落とすことなく自然性を向上することが出来る。

【００２０】請求項４に対応する効果：同じ音素片データを使い振幅だけを変化させると、スピーカーのボリュームを操作したような機械的な変化となり自然性を損なうが、付与すべき振幅に対応する音素片データを選択的に用いることにより、声の大小によって音韻特性に変化を付けることが出来るため、自然音声のダイナミックレンジの広い振幅変化が付与可能となり、より人間の音声に近い合成音声が得られる。

6

【００２１】請求項５に対応する効果：設定すべき音韻継続時間長に対応する音素片データを選択的に用いることにより、音素片の無理な切りつめによる子音の特徴的部分の欠落や、短い母音定常部分の繰り返しによる機械的な音質を避けることが出来、明瞭性を損なうことなく、自然音声のダイナミックレンジの広いテンポ変化が付与可能となり、より人間の音声に近い合成音声を得られる。

【図面の簡単な説明】

【図１】　本発明による音声合成装置の構成を表すブロック図である。

【図２】　図１に示す音素片記憶部のデータ構造を示したものである。

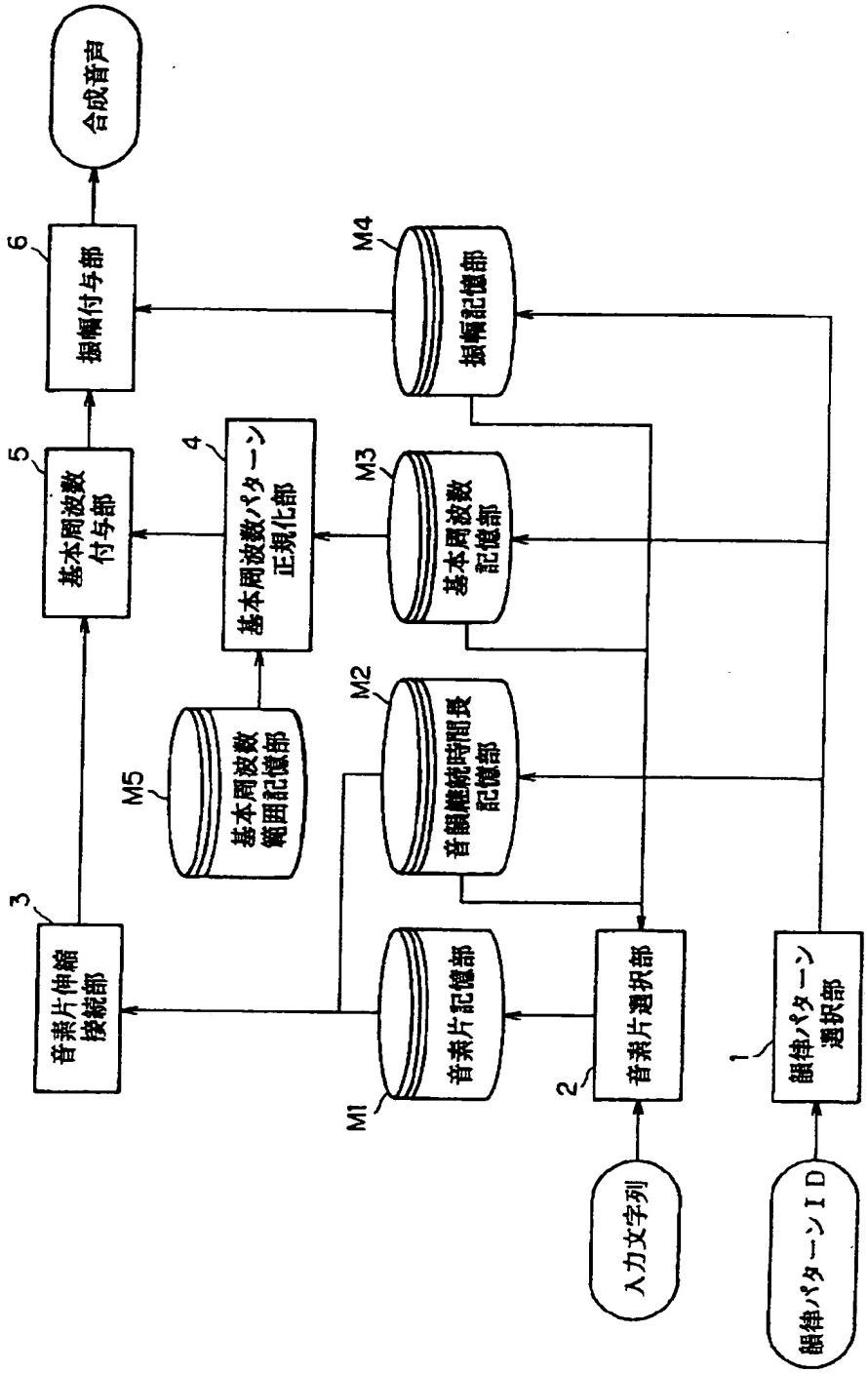
【符号の説明】

１…韻律パターン選択部、２…音素片選択部、３…音素片伸縮接続部、４…基本周波数パターン正規化部、５…基本周波数付与部、６…振幅付与部、Ｍ１…音素片記憶部、Ｍ２…音韻継続時間長記憶部、Ｍ３…基本周波数記憶部、Ｍ４…振幅記憶部、Ｍ５…基本周波数範囲記憶部。

【図２】

ラベル	時間長範囲	基本周波数範囲	振幅範囲	データ
ア	長	50Hz～150Hz	大	ア１
ア	短	100Hz～150Hz	大	ア２
ア	標準	50Hz～100Hz	大	ア３
ア	短	80Hz～150Hz	小	ア４
ア	長	30Hz～100Hz	小	ア５
ア	短	30Hz～80Hz	小	ア６
イ	長	50Hz～120Hz	大	イ１
・・・	・・・	・・・	・・・	・・・

【図１】


合成音声
6
振幅付与部
M4
振幅記憶部
5
基本周波数付与部
4
基本周波数パターン正規化部
M3
基本周波数記憶部
M2
音韻継続時間長記憶部
M5
基本周波数範囲記憶部
3
音素片伸縮接続部
M1
音素片記憶部
2
音素片選択部
1
韻律パターン選択部
入力文字列
韻律パターンＩＤ